Journal of Statistical Physics, Vol. 126, No. 1, January 2007 (© 2006)
DOI: 10.1007/s10955-006-9165-0

Testing Transition State Theory on Kac-Zwanzig Model

G. Ariel[1] and E. Vanden-Eijnden[1]

Received November 22, 2005; accepted July 9, 2006
Published Online: January 3, 2007

A variant of the Kac-Zwanzig model is used to test the prediction of transition state theory (TST) and variational transition state theory (VTST). The model describes the evolution of a distinguished particle moving in a double-well external potential and coupled to N free particles through linear springs. While the Kac-Zwanzig model is deterministic, under appropriate choice of the model parameters the evolution of the distinguished particle can be approximated by a two-state Markov chain whose transition rate constants can be computed exactly in suitable limit. Here, these transition rate constants are compared with the predictions of TST and VTST. It is shown that the application of TST with a naive (albeit natural) choice of dividing surface leads to the wrong prediction of the transition rate constants. This is due to crossings of the dividing surface that do not correspond to actual transition events. However, optimizing over the dividing surface within VTST allows one to eliminate completely these spurious crossings, and therefore derive the correct transition rate constants for the model. The reasons why VTST is successful in this model are discussed, which allows one to speculate on the reliability of VTST in more complicated systems.

KEY WORDS: heat bath, stochastic equation, effective dynamics, harmonic oscillators, transition state theory, metastability, transition rates

1. INTRODUCTION

Deterministic dynamical systems often display very complicated chaotic behavior when the number of degrees of freedom in the systems is large. Amid the complexity of individual trajectories, it is sometimes the case that these trajectories remain confined for very long periods of time in well separated regions of phase-space and only switch from one region to another occasionally. The confinement

[1] Courant Institute of Mathematical Sciences, University, New York, NY 10012, USA; e-mail: ariel@cims.nyu.edu, eve2@cims.nyu.edu

is due to the presence of dynamical bottlenecks between these regions. The system is then said to display metastability, and the regions in which the trajectories remain confined are referred to as metastable sets. Example of systems displaying metastability abound in nature, with examples arising from physics (phase transitions), chemistry (chemical reactions, conformation changes of bio-molecules), biology (regulatory gene networks) and many others. In these systems, it is reasonable to expect that the dynamics can be approximated by a Markov chain over the state-space of the metastable sets with appropriate rate constants. The main questions of interest are determining transition pathways and rates and verifying that the resulting Markov chain does indeed approximate well the dynamics in the system. Unfortunately, these questions are usually highly nontrivial due to the complexity of transition pathways.

One of the earliest attempts to determine transition pathways and rate constants is transition state theory (TST).(13,22,47) TST works under the assumption that the dynamics of the system is ergodic with respect to some known equilibrium distribution and the theory gives the exact average frequency at which trajectories cross a given hypersurface or hyperplane which separates two metastable sets of interest. This average frequency can be used as a first approximation for the frequency of transition between the metastable sets. Unfortunately, it was recognized early on that this approximation can be quite poor, for not every crossing of the dividing surface corresponds to a transition between the metastable sets. Indeed, the trajectories can cross the dividing surface many times in the course of one transition. As a result, the TST prediction for the frequency of transition always overestimates the actual frequency, sometimes grossly so. One way to minimize this problem is to use the freedom in the choice of dividing surface. The best prediction for the frequency from TST is the one corresponding to the dividing surface with minimum crossing frequency. This idea is at the core of the so-called variational transition state theory (VTST),(22,30,41,44) which aims to identify the dividing surface with minimum crossing rate that is the lift-up in phase space of a surface defined in configuration space.

Unfortunately, VTST (just like TST) is an uncontrolled approximation, for it only provides an upper bound on the transition frequency between the metastable sets. In general, one does not know how sharp this bound is. Other assumptions beyond TST and VTST are usually difficult to assess too. Are successive transitions between the metastable sets well approximated by Poisson events (i.e. statistically independent and with exponentially distributed waiting times) as required for the approximation of the dynamics by a Markov chain to hold? How does this property depend on the definition of the metastable sets? Etc.

1.1. The Kac-Zwanzig Model

In this paper we study a benchmark problem, which is, on one hand, simple enough so that many of the assumptions and approximations underlying TST and

VTST can be examined. On the other hand, the model is complex enough to display a wide variety of phenomena common to many dynamical systems exhibiting metastability. The problem we consider is a variant of a model originally proposed by Ford, Kac and Mazur(14,15) and Zwanzig.(48) It was revisited in the context of transition rates in Refs. 7, 20, 31–34 and more recently, from a more analytical point of view in Refs. 1, 5, 19, 21, 23, 25–27, 36.

The Kac-Zwanzig model is a system describing the evolution of a distinguished particle with unit mass and position x_0, moving in an external potential $V(x_0)$ and weakly coupled by a harmonic potential to a bath of N particles of mass $m_i > 0$ with positions x_i, $i = 1, \dots, N$. The Hamiltonian of the system is

$$H(\mathbf{x}, \mathbf{p}) = \frac{1}{2}p_0^2 + V(x_0) + \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + \frac{\gamma}{2N} \sum_{i=1}^{N} (x_i - x_0)^2, \tag{1.1}$$

where, $\gamma > 0$ is a coupling constant. Note that the interaction between the distinguished particle and each bath oscillator is weak and scales as N^{-1}. For short hand we use the vector notation $\mathbf{x} = (x_0, x_1, \dots, x_N)$, $\mathbf{p} = (p_0, p_1, \dots, p_N)$, and p_i is the momentum associated with x_i. The governing equations of motion are:

$$\begin{cases} \ddot{x}_0 = -V'(x_0) - \frac{\gamma}{N} \sum_{i=1}^{N} (x_0 - x_i) \\ \ddot{x}_i = \omega_i^2 (x_0 - x_i) \end{cases} \tag{1.2}$$

where

$$\omega_i^2 = \frac{\gamma}{N m_i}. \tag{1.3}$$

and we will assume that the external potential $V(x_0)$ is the double-well potential

$$V(x_0) = \left(1 - x_0^2\right)^2. \tag{1.4}$$

We also assume that the frequencies $\{\omega_i\}_{i=1,\dots,N}$ are independent and identically distributed (i.i.d.) random variables with probability density function

$$p(\omega) = \begin{cases} \dfrac{2\omega_*}{\pi} \dfrac{1}{\omega_*^2 + \omega^2} & \text{if } \omega \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{1.5}$$

where $\omega_* > 0$ is a parameter playing the role of a characteristic frequency. Unless stated otherwise, we will take $\omega_* = 1$. Notice that all the moments of (1.5) are infinite, i.e. $\omega_* \neq \mathbb{E}\omega_i = \infty$, where $\mathbb{E}$ denotes expectation with respect to (1.5).

The solution of (1.2) lies on the constant energy shell $H(\mathbf{x}, \mathbf{p}) = E$, where E is determined by the initial condition, $E = H(\mathbf{x}(0), \mathbf{p}(0))$. If we assume that the initial condition is such that the energy scales as

$$E = N/\beta, \tag{1.6}$$

for some $\beta > 0$ playing the role of an inverse temperature (i.e., the total energy is an extensive variable), then, in the limit $N \to \infty$ the evolution of the distinguished particle can be described by the following set of stochastic differential equations:[(19,36)]

$$\begin{cases} \dot{x}_0 = p_0 \\ \dot{p}_0 = \sqrt{\gamma} s - V'(x_0) \\ \dot{s} = -s - \sqrt{\gamma} p_0 + \sqrt{2\beta^{-1}}\dot{W}(t) \end{cases} \tag{1.7}$$

where $W(t)$ is a standard Brownian motion. For the reader convenience, this equation is derived in Appendix A.

The existence of the limiting Eq. (1.7) is the main reason why we choose the Kac-Zwanzig model as a test case for TST. Indeed, (1.7) defines a Markov process whose properties can be analyzed via spectral decomposition of the backward operator associated with (1.7). Denoting the eigenvalues by $\lambda_0, \lambda_1, \ldots$, and ordering them as $0 = \lambda_0 < |\lambda_1| < |\lambda_2| < \ldots$, the first zero eigenvalue $\lambda_0 = 0$ is associated with the equilibrium probability distribution of (1.7), whose density is

$$\rho(x_0, p_0, s) = Z^{-1} e^{-\beta(V(x_0) + \frac{1}{2}p_0^2 + \frac{1}{2}s^2)}, \tag{1.8}$$

where $Z = \int_{\mathbb{R}^3} e^{-\beta(V(x_0)+\frac{1}{2}p_0^2+\frac{1}{2}s^2)} dx_0 dp_0 ds$ is a normalization constant. When $\beta \gg 1$, the equilibrium probability distribution is concentrated in small sets around the points $(x_0, p_0, s) = (1, 0, 0)$ and $(x_0, p_0, s) = (-1, 0, 0)$. These sets are metastable, in the sense that any solution of (1.7) spends most of the time inside one of them. Yet, by ergodicity, the solution must hop infinitely often from one set to the other. To understand how these hopping events occur, one must look at the higher eigenvalues of the backward operator associated with (1.7). In the appendix we show that $|\lambda_1| \ll \mathrm{Re}\lambda_2$. This spectral gap indicates that the dynamics in (1.7) (and, hence, also the original one in (1.2) provided that N is large enough and $1 \ll \beta \ll N$, since we took $N \to \infty$ first to arrive at (1.7)) can be approximated by a two-states Markov chain with rates $\frac{1}{2}|\lambda_1|$. In Appendix B we calculate this eigenvalue and its corresponding eigenfunction using the method of matched asymptotics. We show that

$$\lambda_1 = -\frac{|L|}{\sqrt{2}\pi} e^{-\beta}, \tag{1.9}$$

where L is the negative root of the polynomial

$$L^3 - L^2 - (4 - \gamma)L + 4 = 0. \tag{1.10}$$

1.2. Objectives and Organization

The values for the transition rate constant obtained from the limiting Eq. (1.7) can be compared to the predictions of TST and VTST applied to the original Kac-Zwanzig model (1.2). This comparison is the main objective of this paper. Since TST computes the exact rate of crossing of a dividing surface, establishing applicability of TST amounts to answering the following question: can one find a dividing surface such that successive crossings of this surface are statistically independent and exponentially distributed? Here, we will show that the application of TST with a naive (but natural) choice of dividing surface based only on the position of the distinguished particle leads to a wrong prediction for the transition rate constants. This is because the naive dividing surface is crossed many times in the course of each transition between the two metastable sets. However, we will also show that if one optimizes over the dividing surface following VTST, all these spurious crossings can be eliminated completely. Hence, the correct transition rate constants for the model can be computed within VTST. The optimal dividing surface which allows one to do so is then a plane whose normal spans all the configurational degrees of freedom in the system and not only the one associated with the distinguished particle. We shall try to explain why this is the case and when a similar success of VTST can be expected in other more realistic systems.

We note that in Ref. 31, Pollak *et al.* approximate a generalized Langevin equation, which has the same form of the limiting equation in our case, by the Hamiltonian dynamics of the Kac-Zwanzig model, and then use TST for obtaining the escape rates. In Refs. 33 and 34, they obtain the same rates from the limiting equation by a generalization of Kramers method.[18] Although our results are similar, the point of view is different. Here, the Kac-Zwanzig model is used as a platform for analyzing the predictions of TST and VTST and testing the underlying assumptions of these theories. Our results are also all derived from basic principles, and rely on the only (uncontrolled) assumption of ergodicity.

The remainder of this paper is organized as follows. In Sec. 2 we recall the main equilibrium statistical properties of the Kac-Zwanzig model and use them to define two metastable sets for this system. In Sec. 3 we perform a series of detailed numerical experiments to confirm the properties of the system obtained in Sec. 2. In Sec. 4 we develop TST and VTST and find the predictions for the transition rates from these theories. Finally, in Sec. 5 we summarize our findings and discuss possible generalizations. For the reader convenience, we also recall in Appendix A the derivation of the effective stochastic differential equation that describes the dynamics of the distinguished particle in the limit $N \to \infty$ and in Appendix B we calculate the transition rates for the limiting dynamics via asymptotic analysis of the spectrum of the backward operator associated with the effective stochastic differential equations.

2. METASTABILITY IN KAC-ZWANZIG MODEL

Before applying transition state theory to the Kac-Zwanzig model, we must determine over which sets this model is metastable. From the equilibrium probability density function (1.8) of the limiting Eq. (1.7), we know that these sets, once projected onto (x_0, p_0), should reduce to small neighborhood around $(x_0, p_0) = (\pm 1, 0)$. However, here we aim at defining these sets in the original state-space $(\mathbf{x}, \mathbf{p})$. We do so in this section, using the equilibrium statistical mechanics properties of (1.2) which we recall first.

When $1 \ll \beta \ll N$, we have that $E = N/\beta \gg 1$, and the energy shell $H(\mathbf{x}, \mathbf{p}) = E$ is simply connected. We will assume that the dynamics in (1.2) is ergodic on this energy shell, in which case it follows from Birkoff ergodic theorem that time averages can be replaced by ensemble averages with respect to the appropriate equilibrium distribution. We will also assume that this equilibrium distribution is the microcanonical distribution on the energy shell $H(\mathbf{x}, \mathbf{p}) = E$. In other words, we assume that for any suitable test function $g : \mathbb{R}^{N+1} \times \mathbb{R}^{N+1} \to \mathbb{R}$ and almost every initial condition, we have

$$\frac{1}{T}\int_0^T g(\mathbf{x}(t), \mathbf{p}(t))\, dt \to \int_{H(\mathbf{x},\mathbf{p})=E} g(\mathbf{x}, \mathbf{p}) d\mu_E(\mathbf{x}, \mathbf{p}) \quad \text{as } T \to \infty. \tag{2.11}$$

Here μ_E is the microcanonical distribution on $H(\mathbf{x}, \mathbf{p}) = E$,

$$d\mu_E(\mathbf{x}, \mathbf{p}) = \mathcal{Z}^{-1}(E)\frac{d\sigma(\mathbf{x}, \mathbf{p})}{|\nabla H(\mathbf{x}, \mathbf{p})|}. \tag{2.12}$$

where $|\cdot|$ is the standard Euclidean norm in $\mathbb{R}^{2N+2}$, $d\sigma(\mathbf{x}, \mathbf{p})$ denotes a surface element (Lebesgue measure) on $H(\mathbf{x}, \mathbf{p}) = E$, and $\mathcal{Z}(E)$ is a normalization constant. In this paper, the ergodic assumption is not proven rigorously for the potential in (1.4), but it will be corroborated by the numerical experiments presented in Sec. 3.[2]

Assuming ergodicity, let us now show that when the inverse temperature β and the size of the bath N are large enough and satisfy $1 \ll \beta \ll N$, (1.2) displays metastability in the sense that one can find two disjoint sets in phase-space which concentrate most of the probability. These two metastable sets must be regions in phase-space around the minima $\mathbf{x} = \pm(1, 1, \dots, 1)$ of the potential, and as shown below they can be taken as

$$\begin{aligned} S_-(N, \beta, \delta) &= \{(\mathbf{x}, \mathbf{p}) : H(\mathbf{x}, \mathbf{p}) = N/\beta, x_0 < 0, \text{and} \quad H_0(x_0, p_0) < \delta\} \\ S_+(N, \beta, \delta) &= \{(\mathbf{x}, \mathbf{p}) : H(\mathbf{x}, \mathbf{p}) = N/\beta, x_0 > 0, \text{and} \quad H_0(x_0, p_0) < \delta\} \end{aligned} \tag{2.13}$$

[2] Note that in the special case when $V(x_0) = Ax_0^2$, the dynamics in (1.2) is ergodic with respect to the microcanonical equilibrium distribution in (2.12) on every energy shell, provided that the frequencies $A, \omega_1, \dots, \omega_N$ are incommensurable (i.e., linearly independent over the rationals). This is consistent with $\{\omega_i\}_{i=1,\dots,N}$ being i.i.d. drawn from the density in (1.5).

where $H_0(x_0, p_0) = \frac{1}{2}p_0^2 + V(x_0)$ is the energy of the distinguished particle in the absence of the bath, and $\delta \in (0, 1]$ is a parameter. The key reason why the sets $S_\pm$ in (2.13) are metastable is that

$$\lim_{\beta\to\infty}\lim_{N\to\infty}\int_{S_+(N,\beta,\delta)} d\mu_{E=N/\beta}(\mathbf{x},\mathbf{p}) = \frac{1}{2}, \tag{2.14}$$

for every $\delta \in (0, 1]$, and similarly for the integral over $S_-(N, \beta, \delta)$. Note that the order in which the limits are taken matters. Equation (2.14) can be checked by direct calculation. Indeed, performing first the integration over $x_1, \dots, x_N$ and $p_1, \dots, p_N$, we have

$$\int_{S_+(N,\beta,\delta)} d\mu_{E=N/\beta}(\mathbf{x},\mathbf{p}) = \mathcal{Z}_0^{-1}(N/\beta)\int_{H_0<\delta, x_0>0}(1-\beta H_0/N)^{N-1}dx_0dp_0, \tag{2.15}$$

where $\mathcal{Z}_0(N/\beta) = \int_{H_0<N/\beta}(1-\beta H_0/N)^{N-1}dx_0dp_0$ is a normalization constant. In the limit as $N \to \infty$, this is

$$\lim_{N\to\infty}\int_{S_+(N,\beta,\delta)} d\mu_{E=N/\beta}(\mathbf{x},\mathbf{p}) = Z_0^{-1}\int_{H_0<\delta} e^{-\beta H_0}dx_0dp_0, \tag{2.16}$$

where $Z_0 = \int_{\mathbb{R}^2} e^{-\beta H_0}dx_0dp_0$. The Boltzmann-Gibbs probability density function $Z_0^{-1}e^{-\beta H_0}$ is in fact the marginal density for the position and momentum of the distinguished particle in the limit of infinite bath, $N \to \infty$. For every $\delta \in (0, 1]$, each one of the sets $S_\pm$ contains a single energy minimum at $(x_0, p_0) = (\pm 1, 0)$, respectively. Therefore, (2.14) follows from (2.16) by simple evaluation of this integral by Laplace method when $\beta \gg 1$.

We stress that $S_\pm$ are cylindrical sets in $\mathbb{R}^{2N+2}$ around the energy minima. Due to the high dimensionality of the model, the mass of the equilibrium measure is not concentrated in a small volume of phase space. This is another example in which the order of the limits in (2.14) matters.

Equation (2.14) implies that, when $1 \ll \beta \ll N$, any generic trajectory solution of (1.2) spends most of its time in either $S_+(N, \beta, \delta)$ or $S_-(N, \beta, \delta)$. However, under the ergodicity assumption, this trajectory must switch between $S_+(N, \beta, \delta)$ and $S_-(N, \beta, \delta)$ infinitely often. What are the rate constants of these transitions? How do they depend on δ? Are they statistically independent, with transition events Poisson distributed? In other words, can the dynamics in (1.2) be reduced to a Markov process over $S_+(N, \beta, \delta)$ and $S_-(N, \beta, \delta)$ for some suitable choice of δ? These are the questions which we shall investigate in the remainder of this paper, first via a series of numerical experiments with (1.2) (Sec. 3), then within TST and VTST (Sec. 4). Notice that, from the existence and properties of the limiting Eq. (1.7), we know that the dynamics in (1.2) can indeed be reduced to a two-state Markov chain with transition rate (1.9).

3. NUMERICAL EXPERIMENTS

In this section, we perform a series of numerical experiments with (1.2) to investigate when the dynamics can be approximated by a Markov process over the two metastable sets in (2.13). The questions we are especially interested in are:

1. What are the rate of the transition? How do they depend on the parameter γ (interaction strength with the bath) in the model? How do they depend on the choice of δ in (2.13)?
2. Are successive transitions to a good approximation statistically independent? Are the transition times in the sets (2.13) Poisson distributed with intensity equal to the rate of transition? How do these properties depend on δ?

The second question is especially important since it determines when the dynamics in (1.2) can be approximated by a Markov process, and how the metastable sets have to be chosen in this case to get the correct transition rate constants to use in the chain.

For these experiments, we will take $N = 1000$ and $\beta = 7$. We will also consider two different values of γ: $\gamma = 1$ and $\gamma = 10$. In (1.2), the equations of motion of the bath are integrated explicitly, while the equations of motion describing the distinguished particle are integrated numerically using the Verlet algorithm.[45] Each time step is made reversible by a Trotter expansion of the time evolution operator.[17,42] In all the results reported below we use the double-well potential (1.4), but the integration scheme was also checked using the harmonic potential, $V(x_0) = \frac{1}{2}x_0^2$, for which (1.2) can be integrated analytically. Initial conditions are chosen once from the microcanonical invariant distribution on the energy shell $E = N/\beta$. The integration is performed up to time $T = 2 \times 10^6$. The parameters N, T and the step size are chosen so that further increase (in N and T) or decrease (in step size) does not change the average rates considerably. Transition rates are obtained by counting the number of times the trajectory switches between S_+ and S_-. Table I details results obtained for $\delta = 1$ and $\delta = 0.1$. The value $\delta = 0.1$ is arbitrary. Any choice of $0.05 < \delta < 0.8$ yields practically the same rates.

Table I. Transition rates between S_- and S_+ for different values of δ and γ, as obtained in a numerical solution of the full equations of motion (1.2). The number of bath particles is $N = 1000$ and the inverse temperature is $\beta = 7$.

γ	$\delta = 1$	$\delta = 0.1$
1	4.1×10^{-4}	3.7×10^{-4}
10	3.5×10^{-4}	1.1×10^{-4}

The table clearly shows that for large γ, the case $\delta = 1$ is significantly different than with smaller δ.

There are two main sources of errors in the simulation. The first is due to the numerical integrator, and it can be controlled by changing the step size. The second error is, as mentioned above, due to the finite value of the total integration time T. The origin of this error is the theoretical variance in the residence times t_A. This error dominates and it can be evaluated by considering different values of T and by block averaging.(17) Using these techniques, we have estimated the accuracy of the simulation as about 15% for the $\delta = 1$ case, and about 10% with $\delta = 0.1$.

If the dynamics is to be quasi-Markovian, then transitions between S_+ and S_- should have a Poisson distribution. Denoting by τ_δ the waiting time between successive transitions, we expect that it will have an exponential distribution,

$$P[\tau_\delta \geq s] = e^{-ks}, \tag{3.17}$$

for some rate constant k, which is also the average transition rate. Figure 1 depicts the distribution of the waiting times between transitions, $P[\tau_\delta \geq s]$, on a semi-log plot for the case $\gamma = 10$. For $\delta = 1$, the graph is not linear near the origin. However, for $\delta = 0.1$, the linear fit is very good, indicating that transitions events have a Poisson distribution. The slope of the fit is 1.1×10^{-4}, which is also the average transition rate. In order to test the independence of consecutive transitions, we examine the joint distribution of successive waiting times, $\tau_\delta^{(1)}$ and $\tau_\delta^{(2)}$ for various δ and verify that they are statistically independent for δ small enough, and correlated when $\delta = 1$. In particular, we obtain that

$$\frac{\left\langle \tau_{0.1}^{(1)} \tau_{0.1}^{(2)} \right\rangle}{\langle \tau_{0.1} \rangle^2} = 1.08 \qquad \frac{\left\langle \tau_1^{(1)} \tau_1^{(2)} \right\rangle}{\langle \tau_1 \rangle^2} = 0.81 \tag{3.18}$$

where $\langle \cdot \rangle$ denotes a running average over successive τ_δ. In Fig. 2 we also shows that with $\delta = 0.1$, $P[\tau_{0.1}^{(1)} + \tau_{0.1}^{(2)} \geq s] = e^{-ks/2}$, as expected if these times are Poisson distributed and statistically independent.

The main difficulty with the $\delta = 1$ sets is that S_+ and S_- are not well separated (i.e., their closure is not disjoint). A trajectory that reaches the saddle point at $x_0 = 0$ is likely to oscillate between S_+ and S_- several times before completing the transition. Due to such crossing, successive transitions are not independent. This accounts for the jump at $s = 0$ that can be observed in Fig. 1a. With smaller values of δ, correlated crossings are much less frequent and the dynamics can be approximated by a Markov process.

It is also interesting to compare these results with ones obtained using different simulation strategies. For instance, one can redraw new positions and momenta (with the same total energy E) every fixed time segment S, and repeat this procedure T/S times. This strategy is less prone to energy dissipation. Using this

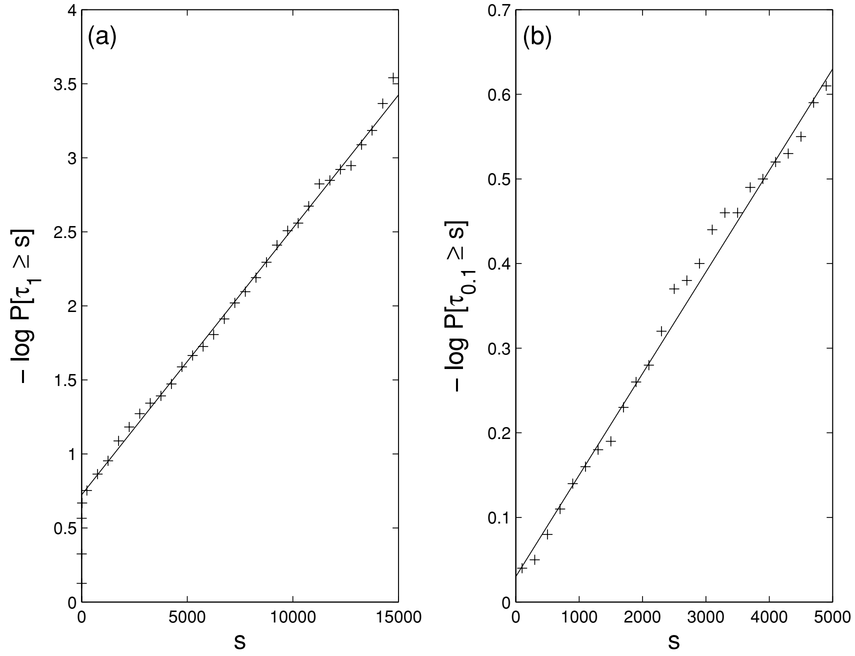

Fig. 1. A semi-log plot of the distribution of the waiting times between transitions, $P[\tau_\delta \geq s]$, for the case $\gamma = 10$. In (a), the sharp jump near the origin is due to rapid re-crossings of the $\{x_0 = 0\}$ plane. In (b), the statistics of transition times confirms the quasi-Markov hypothesis. The slope of the curve is 1.1×10^{-4}, the same as the average transition rate. The graph diverges from the linear fit near the origin.

method, the path traced by $(\mathbf{x}(t), \mathbf{y}(t))$ consists of broken trajectories. However, it still samples the phase space with uniform density on the same energy shell $H(\mathbf{x}, \mathbf{p}) = E$. Ergodicity implies that the two schemes should yield the same rate k for large enough N, T and S. Using this scheme one may also draw initial conditions using the canonical ensemble. The equivalence between the ensembles suggests that averages should be the same for large enough N. Indeed, all schemes yield the same transition rates between the metastable sets (up to the simulation errors).

4. TRANSITION STATE THEORY

In this section we discuss the results of transition state theory (TST)(2,6,47) and variational transition state theory (VTST).(22,30,41) A recent survey of the theory can be found in Ref. 44.

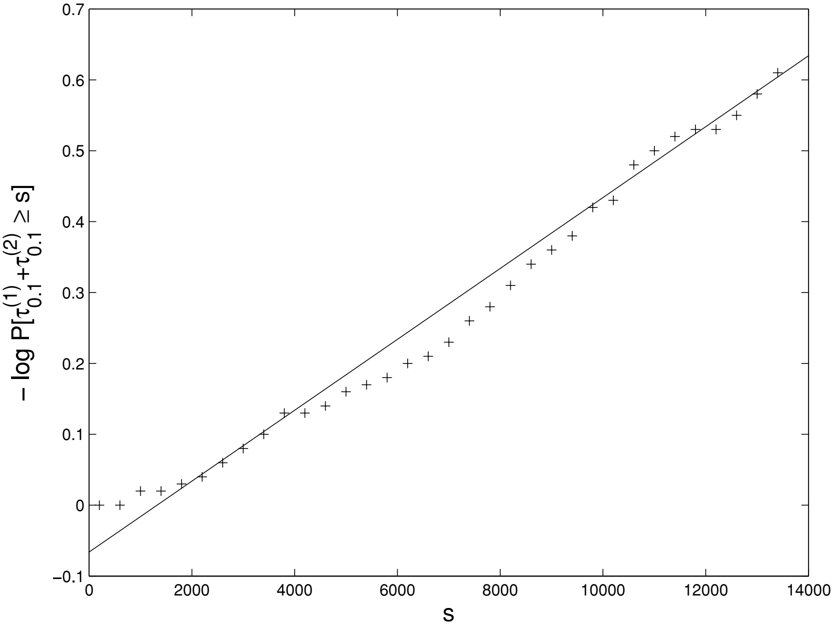

Fig. 2. A semi-log plot of the joint distribution of successive waiting time, $P[\tau_{0.1}^{(1)} + \tau_{0.1}^{(2)} \geq s]$, for the case $\gamma = 10$. As expected, the slope of the curve is 5×10^{-5} which is about half the average transition rate.

The basic idea underling TST is to evaluate the number of transitions by counting the number of times a typical trajectory crosses a hypersurface, or hyperplane, that separates the metastable sets. We shall see that TST can give the exact rate of transition between the two sets (2.13) when $\delta = 1$, and this will correspond to taking $x_0 = 0$ as a dividing surface in the theory. But we also know from Sec. 3 that this rate constant is not the correct one, in the sense that it is not the rate of the two-state Markov chain approximating the dynamics. Recall from Sec. 3 that the correct rate is obtained for sufficiently small values of δ ($\delta \leq 0.8$ was found to be enough), so that the sets S_+ and S_- are well separated in phase-space. So it is far from obvious whether this rate can be obtained within VTST which always divides phase-space into two adjacent regions and gives the rate of transition between these two regions. Nevertheless, we will see that it is indeed so in the Kac-Zwanzig model: the correct rate can be obtained within VTST by optimizing the dividing surface among surfaces whose normal spans only the configurational (position) degrees of freedom. In fact, we will see that this is the case even when the dividing surface is an hyperplane and we will explain how this is possible.

4.1. TST Rate Constant

We define the mean residence time in a subset A in configuration space as the average time a typical trajectory $\mathbf{x}(t)$ spends in A when it visits this set. It is given by

$$t_A = \lim_{T\to\infty} \frac{2}{N_A(T)} \int_0^T \chi_A(\mathbf{x}(t))\,dt, \tag{4.19}$$

where $\chi_A(\mathbf{x})$ is the indicator function of the set A, and $N_A(T)$ is the number of times the trajectory crosses ∂A up to time T:

$$N_A(T) = \int_0^T |\dot{\chi}_A(\mathbf{x}(t))|\,dt. \tag{4.20}$$

The transition rate out of the set A is therefore

$$k_A^{\mathrm{TST}} \equiv \frac{1}{t_A} = \frac{1}{2} \lim_{T\to\infty} \frac{N_A(T)/T}{\frac{1}{T}\int_0^T \chi_A(\mathbf{x}(t))\,dt} = \frac{1}{2} \frac{\lim_{T\to\infty} \frac{1}{T}\int_0^T |\dot{\chi}_A(\mathbf{x}(t))|\,dt}{\lim_{T\to\infty} \frac{1}{T}\int_0^T \chi_A(\mathbf{x}(t))\,dt}. \tag{4.21}$$

By the ergodicity assumption, time averages can be replaced by ensemble averages over the equilibrium distribution μ_E. In the case of a dividing hyperplane passing through the origin, we have $A = \{\mathbf{x} : \mathbf{x}\cdot\hat{n} > 0\}$ for some unit vector $\hat{n}$, and the denominator in (4.21) is

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \chi_{\{\mathbf{x}\cdot\hat{n}>0\}}(\mathbf{x}(t))\,dt = \int_{\mathbf{x}\cdot\hat{n}>0} d\mu_E = \frac{1}{2}, \tag{4.22}$$

where we used the symmetry of the potential. Similarly, the numerator is

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T |\dot{\chi}_A(\mathbf{x}(t))|\,dt = \int_{\mathbb{R}^{2N+2}} |\hat{\mathbf{n}}\cdot M^{-1}\mathbf{p}|\delta(\hat{\mathbf{n}}\cdot\mathbf{x})\,d\mu_E, \tag{4.23}$$

where M is a diagonal $(N+1)\times(N+1)$ matrix whose entries are the masses, $1, m_1 \ldots, m_N$. Inserting (4.22) and (4.23) in (4.21) we arrive at

$$k_{\hat{n}}^{\mathrm{TST}} = \int_{\mathbb{R}^{2N+2}} |\hat{\mathbf{n}}\cdot M^{-1}\mathbf{p}|\delta(\hat{\mathbf{n}}\cdot\mathbf{x})\,d\mu_E, \tag{4.24}$$

where we have denoted k_A^{TST} by $k_{\hat{n}}^{\mathrm{TST}}$ since the set $A = \{\mathbf{x} : \mathbf{x}\cdot\hat{n} > 0\}$ is determined by $\hat{n}$. To evaluate (4.24) we make a change of variables into mass weighted coordinates

$$\mathbf{y} = M^{1/2}\mathbf{x}, \qquad \mathbf{q} = M^{-1/2}\mathbf{p}, \tag{4.25}$$

to arrive at

$$k_{\hat{n}}^{\mathrm{TST}} = \int_{\mathbb{R}^{2N+2}} |\hat{l}\cdot\mathbf{q}|\delta(\hat{l}\cdot\mathbf{y})\,d\tilde{\mu}_E. \tag{4.26}$$

Here we defined

$$\hat{l} = \frac{M^{-1/2}\hat{n}}{|M^{-1/2}\hat{n}|}, \tag{4.27}$$

and $\tilde{\mu}_E$ is the microcanonical distribution obtained by writing the Hamiltonian (1.1) in the new coordinates system:

$$\begin{aligned} H(\mathbf{y}, \mathbf{q}) &= \frac{1}{2}q_0^2 + \frac{1}{2}\sum_{i=1}^{N} q_i^2 + U(\mathbf{y}) \\ U(\mathbf{y}) &= V(y_0) + \frac{\gamma}{2N}\sum \left(m_i^{-1/2} y_i - y_0\right)^2. \end{aligned} \tag{4.28}$$

For fixed $\mathbf{y}$, the energy shell $H(\mathbf{y}, \mathbf{q}) = E$ is an $N+1$ dimensional sphere. Integrating over the new momenta coordinates yields

$$\begin{aligned} k_{\hat{n}}^{\text{TST}} &= C_N \int_{U(\mathbf{y})<N/\beta} \delta(\hat{l}\cdot\mathbf{y}) \left(1 - \frac{\beta}{N}U(\mathbf{y})\right)^{N/2} d\mathbf{y} \\ &= C_N \int_{U(\mathbf{y})<N/\beta, \hat{l}\cdot\mathbf{y}=0} \left(1 - \frac{\beta}{N}U(\mathbf{y})\right)^{N/2} d\sigma, \end{aligned} \tag{4.29}$$

where in the second equality we have changed the $N+1$ dimensional volume integration over the Dirac delta distribution to a N dimensional surface integral. The constant C_N is obtained by properly accounting for the normalization of the distribution $\tilde{\mu}_{E=N/\beta}$. A calculation similar to the one that led to (2.14) gives

$$C_N \sim \frac{1}{2}(2\pi)^{-N/2+1}\left(\frac{N}{\beta}\right)^{-N}, \tag{4.30}$$

where the asymptotic equality sign $\sim$ means that ratio of the expressions on both side tends to 1 as $N \to \infty$, $\beta \to \infty$ (in this order).

At this point we will assume that the external potential $V(y_0)$ in $U(\mathbf{y})$ can be expanded to second order in y_0 around $y_0 = 0$. This approximation is justified at the end of this subsection where it is shown to be valid when the minimum of $U(\mathbf{y})$ on the plane $\{\mathbf{x}\cdot\hat{n} = 0\}$ is attained at the origin. When this is the case higher order terms introduce a correction of the order of $1/\beta$ in the limit as $N \to \infty$, $\beta \to \infty$ with $\beta/N \to 0$. Expanding the potential $V(y_0)$ to second order in y_0 yields

$$k_{\hat{n}}^{\text{TST}} = C_N \int_{U_{\text{quad}}(\mathbf{y})<N/\beta, \hat{l}\cdot\mathbf{y}=0} \left(1 - \frac{\beta}{N}U_{\text{quad}}(\mathbf{y})\right)^{N/2} d\sigma, \tag{4.31}$$

where

$$U_{\text{quad}}(M^{-1/2}\mathbf{y}) = 1 - 2y_0^2 + \frac{\gamma}{2N}\sum_{i=1}^{N}\left(m_i^{-1/2}y_i - y_0\right)^2, \tag{4.32}$$

The integral in (4.31) can be straightforwardly evaluated in the limit as $N \to \infty$, $\beta \to \infty$ with $\beta/N \to 0$. This gives

$$k_{\hat{n}}^{\mathrm{TST}} \sim \frac{\sqrt{2}}{\pi}\left(\Pi_{i=1}^{N} m_i\right)^{-1/2}\left(\frac{\gamma}{N}\right)^{N/2}\left(|\det H_N^n|\right)^{-1/2} e^{-\beta} \tag{4.33}$$

where $H_N^{\hat{n}}$ denote the Hessian obtained by restricting $U_{\mathrm{quad}}(\mathbf{y})$ to the N dimensional linear subspace perpendicular to $\hat{n}$.

Justification of (4.33). Consider the integral in (4.29):

$$I = \int_{U(\mathbf{y})<N/\beta} \delta(\hat{l}\cdot\mathbf{y})\left(1-\frac{\beta}{N}U(\mathbf{y})\right)^{N/2} d\mathbf{y}. \tag{4.34}$$

We show that on the plane $\hat{l}\cdot\mathbf{y}$, the situation is similar to the case of (2.14), and the restricted potential can be approximated by a quadratic expansion in y_0. In order to see that, assume, without loss of generality, that $l_N \neq 0$. Integrating out y_N yields

$$I = \int_{\bar{U}(\mathbf{y})<N/\beta} \left(1-\frac{\beta}{N}\bar{U}(\mathbf{y})\right)^{N/2} d\mathbf{y}, \tag{4.35}$$

where

$$\bar{U}(\mathbf{y}) = V(y_0) + \frac{\gamma}{2N}\sum_{i=1}^{N-1}\left(m_i^{-1/2}y_i - y_0\right)^2 + \frac{\gamma}{2N}\left(m_N^{-1/2}l_N^{-1}\sum_{i=0}^{N-1} l_i y_i + y_0\right)^2. \tag{4.36}$$

This potential has the form

$$\bar{U}(\mathbf{y}) = V(y_0) + C(y_0) + \langle \mathbf{y}' - b(y_0), A(\mathbf{y}' - b(y_0))\rangle, \tag{4.37}$$

where $C(y_0)$ is a quadratic function of y_0, $\mathbf{y}' = (y_1, \ldots, y_{N-1})$, A is a $(N-1)\times(N-1)$ constant, positive definite matrix and $b(y_0)$ is a vector in $\mathbb{R}^{N-1}$, which is linear in y_0. Integrating out $y_1, \ldots, y_N$ yields

$$I = D_N(\det A)^{-1/2}\int_{V(y_0)+C(y_0)<N/\beta}\left(1-\frac{\beta}{N}(V(y_0)+C(y_0))\right)^{N-1/2} dy_0, \tag{4.38}$$

with

$$D_N = \frac{\sqrt{\pi}}{2^N}\left(\frac{N}{\beta}\right)^{(N-1)/2}\frac{N\Gamma(N-1)S_{N-1}}{\Gamma(N+1/2)}, \tag{4.39}$$

where $\Gamma(z)$ denotes the Euler Gamma function. For fixed β, the integrand of (4.38) converges uniformly to an exponent in the limit $N \to \infty$. Hence,

$$I \sim D_N(\det A)^{-1/2} \int_{\mathbb{R}} e^{-\beta(V(y_0)+C(y_0))} dy_0, \tag{4.40}$$

and we can use the Laplace approximation, which amounts to expanding $V(y_0) + C(y_0)$ to second order in y_0. Since $C(y_0)$ is quadratic, this is the same as expanding $V(y_0)$.

4.2. Naive TST: Rate Across the Plane $x_0 = 0$

Since the metastability originates from the double-well potential, a naive (but at this point not so natural anymore) candidate for a dividing surface is the plane $\{x_0 = 0\}$. Substituting $\hat{n} = \hat{l} = (1, 0, \ldots, 0)$ into (4.33) yields,

$$k^{\text{TST}}_{\{x_0=0\}} \sim \frac{\sqrt{2}}{\pi} e^{-\beta}. \tag{4.41}$$

This calculation can also be easily derived from the first line of (4.29) by integrating out y_0 and eliminating the delta function. The factor in the exponential is β times the energy barrier, which was taken to be one. Note there is no dependence in γ, which is in clear contradiction to the numerical results for the transition rates with $\delta = 0.1$. Taking $\beta = 7$, the value used for the numerical simulation yields $k^{\text{TST}}_{\{x_0=0\}} = 4.1 \times 10^{-4}$, which is in very good agreement with the numerical result for $\delta = 1$. This is expected because with $\delta = 1$, the metastable sets S_+ and S_- are tangent to the plane $\{x_0 = 0\}$.

4.3. VTST: Rate Across the Plane with Minimum Recrossing

Due to the possible recrossings of the dividing, TST always over counts the number of transitions between the metastable sets $S_\pm$, especially in the relevant case when these sets are separated in phase space. One way to improve this result is to minimize the TST rate for different choices of hypersurfaces. In variational TST (VTST) we optimize the rate across all the dividing surfaces that are the lift-up in phase space of dividing surfaces in configuration space. In our application of VTST, we shall further assume that the dividing surfaces are hyperplanes which satisfy the condition that the minimum of the potential energy $U(\mathbf{y})$ is obtained at the origin. At the end of this Section we show that the VTST prediction for the transition rates is in this case given by

$$k^{\text{VTST}} = \frac{|L|}{\sqrt{2}\pi} e^{-\beta} \tag{4.42}$$

Table II. A comparison between transition rates obtained by TST, VTST and simulation. Transition rates between S_- and S_+ for different values of δ. The TST column shows the TST predictions with the $\{x_0 = 0\}$ plane, given by Eq. (4.41). The VTST column shows the Variational TST prediction (4.42). Simulation results are obtained with $N = 1000$ bath particles and inverse temperature is $\beta = 7$.

γ	TST $\{x_0 = 0\}$	sim $\delta = 1$	VTST	sim $\delta = 0.1$
1	4.1×10^{-4}	$4.1 \times 10^{-4} \pm 4 \times 10^{-5}$	3.8×10^{-4}	$3.7 \times 10^{-4} \pm 2 \times 10^{-5}$
10	4.1×10^{-4}	$3.5 \times 10^{-4} \pm 4 \times 10^{-5}$	1.2×10^{-4}	$1.1 \times 10^{-4} \pm 1 \times 10^{-5}$

where L is the unique positive solution of the cubic equation

$$L^3 + L^2 - (4 - \gamma)L - 4 = 0. \tag{4.43}$$

This is exactly the same as the limiting rate (1.10). Equation (4.43) was previously obtained by Pollak *et al.* in Ref. 31. Note that unlike the TST result, the rate (4.42) depends on γ (for large γ, L and hence k^{VTST} is inversely proportional to γ). Note also that in our analysis we take the high temperature limit, $\beta \to \infty$, first, and only then $\gamma \to 0$. Switching the order of these limit may change these predictions.[16]

The direction of the plane leading to the VTST rate in (4.42) is random. However, it spans all of the configurational degrees of freedom. Table II compares the predictions of TST and VTST with the numerical simulations.

The error of the theoretical predictions of TST and VTST are of the order of $1/\beta$. Taking these errors into account, it shows that in the Kac-Zwanzig model, VTST gives the correct value for the transitions rate between the set S_+ and S_- even in the relevant case when δ is small and the successive transition are Poisson events. This surprising result will be elucidated in Sec. 4.4. In contrast, the naive TST prediction, which is independent of γ, provides only a rough approximation when δ is small and γ is large.

It is also interesting to note that in our model, the plane separating the two minima points, $\pm(1, 1, \dots, 1)$, that is being crossed the fewest number of times is not a good indication for the transition rate between the metastable states. The lowest rate is obtained on a plane that is perpendicular to the direction corresponding to the bath particle that has the lowest frequency. Since the interaction of a single particle with x_0 is weak (of the order of $1/N$), that particle will simply oscillate close to its natural frequency, which is also of order $1/N$. The minimum of the Hamiltonian on this plane is close to zero for large N, and is obtained away from the origin. This minimum must be excluded because the corresponding dividing plane intersects the metastable sets S_+ and S_-. Adding the constraint that on the plane, the minimum of the potential is obtained at the origin, the minimum rate is the one given in (4.42).

Derivation of (4.42). The full Hessian matrix at zero, H_{N+1}, has a single negative eigenvalue, λ_-, and N positive ones. Therefore, the plane that will maximize the restricted Hessian is perpendicular to the eigenvector that corresponds to λ_-. Therefore, $H_N = H_{N+1}/\lambda_-$. The full Hessian matrix at the origin reads

$$(H_{N+1})_{ij} = \left.\frac{\partial^2 U(\mathbf{y})}{\partial y_j \partial y_k}\right|_{\mathbf{y}=0}, \tag{4.44}$$

where,

$$\begin{aligned} H_{00} &= -4+\gamma \\ H_{0i} &= -\frac{\gamma}{N} m_i^{-1/2} \\ H_{ij} &= \frac{\gamma}{N} m_i^{-1} \delta_{ij}. \end{aligned} \tag{4.45}$$

Performing the row manipulation $\text{row}_0 \to \text{row}_0 + \sqrt{m_k} \cdot \text{row}_k$ for every $k = 1 \ldots N$, the matrix becomes lower triangular. Its determinant is

$$H_{N+1} = -4\left(\frac{\gamma}{N}\right)^N (\Pi_i m_i)^{-1/2}. \tag{4.46}$$

Substituting into (4.33),

$$k_{\hat{n}}^{\mathrm{TST}} \sim \frac{\sqrt{|\lambda_-|}}{\sqrt{2}\pi} e^{-\beta}. \tag{4.47}$$

Denoting by λ and $\mathbf{l} = (1, l_1, \ldots, l_N)$ the eigenvalues and eigenvectors, they satisfy

$$\begin{cases} (\gamma-4) - \frac{\gamma}{N}\sum_{i=1}^N m_i^{-1/2} l_i = \lambda \\ -\frac{\gamma}{N} m_i^{-1/2} + \frac{\gamma}{N} m_i^{-1/2} l_i = \lambda l_i, \quad i = 1, \ldots, N. \end{cases} \tag{4.48}$$

Solving for l_i and substituting back we obtain an equation for λ

$$\gamma - 4 - \frac{\gamma}{N}\sum_{i=1}^N \frac{\omega_i^2}{\omega_i^2 - \lambda} = \lambda, \tag{4.49}$$

where we used the definition of $m_i = \gamma/(N\omega_i^2)$. Looking for the negative eigenvalue, we denote $\lambda = -L^2$. By the strong law of large numbers, as $N \to \infty$, (4.49) reduces to

$$\gamma - 4 - \gamma \mathbb{E}\left[\frac{\omega_1^2}{\omega_1^2 + L^2}\right] = -L^2. \tag{4.50}$$

Using the probability density for the frequencies ω in (1.5), we have for $L > 0$,

$$\mathbb{E}\left[\frac{\omega_1^2}{\omega_1^2 + L^2}\right] = \frac{1}{1+L}. \tag{4.51}$$

Substituting into (4.50), L solves

$$L^3 + L^2 - (4 - \gamma)L - 4 = 0. \tag{4.52}$$

This polynomial has a single positive root for all $\gamma \geq 0$.

4.4. Local Dynamics Around the Hyperbolic Point: Why Does VTST Works While Naive TST Does Not?

Following Refs. 2, 6, 24, it has been proved in Refs. 40, 43, 46 that one can always find a hypersurface in phase-space that separates the metastable sets (S_+ and S_- in our case) and is crossed exactly once in each transition. As shown above in Sec. 4.3, in the Kac-Zwanzig model it is possible to find a hyperplane whose normal spans configuration space with the same properties, which is quite remarkable. We now try to explain why in this model this is the case. The discussion should shed some light on the more general question of when one can find a hyperplane in configuration space that is crossed exactly once between transitions. We also discuss the problem of finding a hypersurface that is not necessarily a plane.

After changing variables to mass weighted coordinates, we can think about the system as a particle with unit mass that is moving in a $N + 1$ dimensional potential. From the discussion justifying the quadratic approximation, we also know that almost all trajectories that go from one metastable set to the other pass with $y_0 \ll 1$. Locally, the potential $U(\mathbf{y})$ is a quadratic form. It is important to note that by locally we means a cylindrical set $\{\mathbf{y} | y_0 < \epsilon\}$, and not a small sphere around the origin. The quadratic form, $\bar{U}(\mathbf{y})$, has N stable directions and a single unstable one, $\hat{l}$. In this coordinate system, the dynamics has the form

$$\begin{cases} \ddot{z}_0 = -\lambda_- z_0 \\ \ddot{z}_i = \lambda_i z_i, \quad i = 1, \ldots, N \end{cases} \tag{4.53}$$

where $z_0 = \mathbf{y} \cdot \hat{l}$, the component of $\mathbf{y}$ in the unstable direction, and $z_1, \ldots, z_N$ correspond to the components of $\mathbf{y}$ in the directions of the remaining (stable) eigenvectors. Hence, $\lambda_-, \lambda_i > 0$. The value of λ_- was obtained in the limit $N \to \infty$, with probability one, in (4.52). In this coordinates system, the $N + 1$ variables are uncoupled. Out of these, $z_1, \ldots, z_N$ are harmonic oscillators with frequencies $\sqrt{\lambda_i}$. Since they are independent random variables, some of the frequencies are very large. The corresponding variables oscillate on a time scale that is much shorter than the dynamics of z_0, determined by the value of λ_-. These considerations imply that if we consider any plane that is not perpendicular to $\hat{l}$, the trajectory of the full system $(z_0, z_1, \ldots, z_N)(t)$ will cross that plane many times while the particle is close to the saddle point. The transition rate of such a plane will be large, making it a poor candidate for TST. On the other hand, the plane perpendicular to $\hat{l}$ is crossed exactly once in every transition between metastable sets. This is because the z_0 variable cannot "turn around" and return to the origin after crossing it. In the VTST calculation, we use the same reasoning for finding the optimal rate. It is important to stress the point that unlike other models,(16) the dynamics at finite N is Hamiltonian. In particular, the origin of spurious recrossings of the TST candidate plane are oscillations in stable directions and not by random noise.

Local consideration such as the one described in the previous paragraph may fail due to two possible factors. The first is due to the quadratic approximation (which also justifies the planar approximation). This issue was addressed in the justification of (4.29). In systems with high dimensions, one cannot assume that the trajectories switching between metastable sets pass close to the origin. In our model, the approximation is justified since all N directions are quadratic to begin with. The second problem has to do with our definition of a transition. A "true" transition is a trajectory that starts at one metastable set and ends in another. Once the particle gets out of the range in which the quadratic approximation holds, we do not know *a priori* whether it moves to the new metastable set or turns around back without completing the transition. In the Kac-Zwanzig model, our numerical experiments indicate that the probability that the particle does not complete the transition is extremely small. This is also confirmed via the analysis of the limiting Eq. (1.7) (see Appendix B). These nonlocal effects will be more difficult to assess in more general systems. An example in which the particle returns to the separating surface before completing the transition is discussed in Ref. 16.

5. CONCLUDING REMARKS

In some way, one can say that TST and VTST provide the exact answer to a question which is usually the wrong one. These theories give the exact average frequency of crossing a dividing surface in phase space that separates two metastable sets of interest. Unfortunately, this exact frequency is not, in general, the actual frequency of crossing between the sets because trajectories may cross the dividing surface many times in the course of each transition between the sets. As a result, the TST frequency is a upper bound, and sometimes a poor one, of the actual frequency one should use in the Markov chain description of the effective dynamics of the system. VTST does better than TST, as it minimizes the rate of crossing among a given class of dividing surfaces. However, it may not give the correct rate either.

In this paper, the difficulties with TST were illustrated on the Kac-Zwanzig model. Quite surprisingly, it was also shown that VTST gives the exact rates of crossing between the two metastable sets in this model. Unfortunately, this nice conclusion should be amended by a word of caution.

The argument presented in Sec. (4.4) is local in nature, both because it uses a quadratic approximation of the potential near the saddle point and because it requires that once a trajectory goes a little bit away from the dividing surface, then it makes it all the way to one of metastable sets before returning to the surface long afterward. These assumptions are justified in the Kac-Zwanzig model, but they could very possibly fail in more complicated systems. Because the transition region is not localized at a saddle point, the VTST dividing surface must in general be a more complicated surface than the lift-up in phase space of a hyperplane in configuration space. In fact, it is likely that in many situations one should use other techniques, such as TPS[3,8] or the string method,[10–12,35] which are more sophisticated than TST and VTST, to obtain a nonlocal description of the transition pathways between the metastable sets. It is nevertheless encouraging that VTST can be used with success at least on some systems, and that the reasons for its success or failure can be understood from the behavior of the trajectories in and out of the dividing surface.

APPENDIX A: DERIVATION OF (1.7)

Here we proceed informally and refer the reader to Ref. 1 for a mathematically rigorous derivation of (1.7). Similar rigorous derivations can be found in Refs. 5, 19, 21, 36. The derivation presented here uses the canonical ensemble rather than the micro-canonical one, i.e., at $t = 0$ bath particles are distributed according to their equilibrium Gibbs measure which is given by

$$\begin{aligned} x_i(0) &= \mathcal{N}(x_0(0), N/\beta\alpha_i) \\ p_i(0) &= \mathcal{N}(0, m_i/\beta), \end{aligned} \tag{A.1}$$

where $\mathcal{N}(m, \sigma^2)$ denotes the Gaussian distribution with mean m and variance σ^2. In Ref. 1 we treat the case of microcanonical initial conditions and prove that the solution of the equation of motion at finite N, (1.2), converges strongly (in L^2) to the solution of the limiting Eq. (1.7) in every finite time segment $[0, T]$, $T < \infty$.

The dynamics of the full system of $N + 1$ particles is given by (1.2). Integrating the bath variables x_i yields

$$\ddot{x}_0 + V'(x_0) + \int_0^t R_N(t-\tau)\dot{x}_0(\tau)d\tau = \frac{1}{\sqrt{\beta}}\xi_N(t), \tag{A.2}$$

where

$$R_N(t) = \frac{\gamma}{N}\sum_{i=1}^{N}\cos\omega_i t, \tag{A.3}$$

and

$$\xi_N(t) = -\sqrt{\frac{\beta\gamma}{N}} \sum_{i=1}^{N} \left([x_i(0) - x_0(0)] \cos \omega_i t + p_i(0) \frac{\sin \omega_i t}{\omega_i m_i} \right). \tag{A.4}$$

Note that initial conditions appear only in ξ_N, which plays the role of a random noise. It is a Gaussian process with zero mean and covariance

$$\mathbb{E}_0 \left[\xi_N(t_1) \xi_N(t_2) \right] = R_N(t_1 - t_2). \tag{A.5}$$

In the limit $N \to \infty$, the strong law of large numbers implies that for any fixed t, $R_N(t)$ will converge to its average $R(t)$

$$R(t) = \lim_{N\to\infty} R_N(t) = \gamma \lim_{N\to\infty} \frac{1}{N} \sum \cos \omega_i t = \gamma \mathbb{E}[\cos \omega t] = \gamma e^{-|t|}. \tag{A.6}$$

In order to evaluate the rate of convergence, we calculate the second moment of R_N. Breaking all double sums into the diagonal and off-diagonal parts yields

$$\mathbb{E}[R_N(t_1) R_N(t_2)] - \mathbb{E}[R_N(t_1)] \mathbb{E}[R_N(t_2)] = O\left(\frac{1}{N}\right). \tag{A.7}$$

We conclude that the limiting equation describing the dynamics of the distinguished particle is

$$\ddot{x}_0 + V'(x_0) + \gamma \int_0^t e^{-(t-\tau)} \dot{x}_0(\tau) d\tau = \frac{1}{\sqrt{\beta}} \xi(t), \tag{A.8}$$

where $\xi(t)$ is a Gaussian process with zero mean and covariance function $\gamma e^{-|t|}$. Hence, it is an Ornstein-Uhlenbeck process at equilibrium which solves the stochastic differential equation

$$d\xi = -\xi\, dt + \sqrt{2\gamma} dW_t, \qquad \xi(0) = \mathcal{N}(0, 1). \tag{A.9}$$

(A.8) and (A.9) can now be written in the form of (1.7) with $s(0) = \xi(0)/\sqrt{\beta\gamma}$. This equation can also be written as

$$\frac{d}{dt} \begin{pmatrix} x_0 \\ p_0 \\ s \end{pmatrix} = -K \nabla \mathcal{H}(x_0, p_0, s) + \sqrt{\frac{2}{\beta}} \sigma \dot{W}_t. \tag{A.10}$$

Here $\mathcal{H}(x_0, p_0, s) = V(x_0) + \frac{1}{2} p_0^2 + \frac{1}{2} s^2$,

$$\sigma = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \tag{A.11}$$

and $K = K^S + K^A$, with

$$K^S = \sigma\sigma^T = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$K^A = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & -\sqrt{\gamma} \\ 0 & \sqrt{\gamma} & 0 \end{pmatrix}. \tag{A.12}$$

This form is used in Appendix B.

APPENDIX B: TRANSITION RATE FOR THE LIMITING EQUATION

In this section we detail the calculation leading to the theoretical transition rate of the effective dynamics of the distinguished particle. Similar calculations in the gradient case can be found in Refs. 29, 37, 38, but to the best of our knowledge the result for non-gradient systems of type (A.10) is new. As the calculation shows, there is exactly one nontrivial eigenvalue that tends to zero as $\beta \to \infty$, and the corresponding eigenvector determines the metastable sets.

We assume the dynamics satisfies an equation of motion in $\mathbb{R}^d$ of the form

$$\frac{du}{dt} = -K\nabla\mathcal{H}(u) + \sqrt{\frac{2}{\beta}}\sigma\dot{W}_t, \tag{B.1}$$

where $\mathcal{H}$ is the Hamiltonian, σ a $d \times m$ matrix such that $\sigma\sigma^T = K^S$, the symmetric part of K, and B_t is an m dimensional BM. We also assume that the eigenvalues of K have positive real part. This implies that the stability of extremum points of $K\nabla\mathcal{H}$ is of the same type as with $\nabla\mathcal{H}$. In the model considered here, these parameters are given by (A.10), hence the equation is in three dimensions.

B.1. The Eigenvalue Problem

The eigenvalue problem associated with (B.1) is

$$L\varphi \equiv -(K\nabla\mathcal{H}) \cdot \nabla\varphi + \frac{1}{\beta}K^S : \nabla\nabla\varphi = -\lambda\varphi \tag{B.2}$$

This equation admits a trivial solution $\varphi_0 = 1$ with $\lambda = 0$, consistent with the existence of a unique equilibrium distribution for (B.1). It is easily verified that this distribution is the Boltzmann-Gibbs distribution with density

$$\psi(u) = Z^{-1}e^{-\beta\mathcal{H}} \qquad \text{where} \quad Z = \int_{\mathbb{R}^d} e^{-\beta\mathcal{H}}du \tag{B.3}$$

The second eigenvalue, $-\lambda_1$, is real and of the order of $\lambda_1 \approx e^{-\beta A}$, as we show next. The analysis also allows one to construct only one such small eigenvalue, i.e. it shows that this is the unique vanishing eigenvalue in the system and $\lambda_1 \ll \operatorname{Re} \lambda_2$, implying bistability.

B.2. Perturbation in β^{-1}

Using regular expansion in β^{-1}, to leading order the eigenvector, φ_1, associated with $-\lambda_1$ satisfies

$$-(K\nabla\mathcal{H}) \cdot \nabla\varphi_1 = 0. \tag{B.4}$$

Hence, φ_1 is constant along the flow lines of $K\nabla\mathcal{H}$. The critical points under the flow of $K\nabla\mathcal{H}$ are the same as that of $\nabla\mathcal{H}$. The origin is a saddle points while $(\pm 1, 0, 0)$ are stable. In the limit of zero temperature, $\beta \to \infty$, the dynamics follows the flow lines of $K\nabla\mathcal{H}$. Therefore, for large β the metastable sets are the two basin of attractions of the stable minima. The stable manifold through the origin is the hypersurface that separates the two sets.

Denote by S the hypersurface separating the two basins of attractions and by $A_\pm$ the regions flowing toward $(\pm 1, 0, 0)$, respectively. The flow of $K\nabla\mathcal{H}$ on S is tangent to the surface, i.e. S is a stable invariant manifold for the dynamics $\dot{u} = -K\nabla\mathcal{H}$. From (B.4), we have

$$\varphi_1 = \begin{cases} C_+ & \text{if } u \in A_+ \\ C_- & \text{if } u \in A_- \end{cases} \tag{B.5}$$

The constants C_+ and C_- can be determined from the orthogonality condition $\int_{\mathbb{R}^n} \varphi_1 \psi \, du = 0$, and the normalization condition $\int_{\mathbb{R}^n} \varphi_1^2 \psi \, du = 1$. To leading order in β^{-1} these are explicitly:

$$\begin{cases} C_+^2 N_+ + C_-^2 N_- = 1 \\ C_+ N_+ + C_- N_- = 0, \end{cases} \tag{B.6}$$

where

$$N_\pm = \int_{A_\pm} \psi \, du. \tag{B.7}$$

Since H is symmetric we have $N_+ = N_- = 1/2$, and without loss of generality, we can take $C_\pm = \pm 1/\sqrt{2}$.

Equation (B.4) fails around S. Therefore, in the vicinity of this surface a boundary layer type of analysis must be performed to determine the behavior of ψ_1 and, eventually, evaluate λ_1. This can be done by writing (B.2) using a local coordinates system around the hypersurface S. Decompose $u = (r, z)$, where z is a local coordinate system on S and r is the signed distance between u and S, counted positively in A_+ and negatively in A_-. For definiteness, we will assume that the saddle point on S is located at $(z, r) = (0, 0)$. Let $\eta = \sqrt{\beta} r$ and look for a solution of (B.2) as a function of (η, z). The Hamiltonian, its derivatives, and φ_1 can then be expanded in power of β^{-1} near the surface S. Denoting by $\hat{n}(z)$ the unit normal to S at z and using $K\nabla\mathcal{H}(0, z) \cdot \hat{n}(z) = 0$, (B.2) becomes, to leading order β^0,

$$a(z)\eta\frac{\partial\varphi_1}{\partial\eta} + b(z)\frac{\partial^2\varphi_1}{\partial\eta^2} - [K\nabla\mathcal{H}(0, z)] \cdot \nabla_z\varphi_1 = 0 \tag{B.8}$$

Here ∇_z denotes the projection of the gradient onto S and

$$\begin{aligned} a(z) &= -\hat{n}(z) \cdot [K\nabla\nabla\mathcal{H}(0, z)\hat{n}(z)] \\ b(z) &= \hat{n}(z) \cdot K\hat{n}(z) \end{aligned} \tag{B.9}$$

Notice that both $a(z) > 0$ and $b(z) > 0$ since S is a stable invariant manifold.

B.3. Solution of (B.8)

Equation (B.8) must be solved with the boundary condition $\lim_{\eta\to\pm\infty} \varphi(\eta, z) = C_\pm$. Look for a solution of (B.8) in the form

$$\varphi_1(\eta, z) = \phi_1(\eta c(z)) \tag{B.10}$$

for some function $c(z) > 0$ to be determined later. Letting $\zeta = \eta c(z)$, (B.8) becomes

$$\tilde{a}(z)\zeta\frac{d\phi_1}{d\zeta} + c^2(z)b(z)\frac{d^2\phi_1}{d\zeta^2} = 0 \tag{B.11}$$

where

$$\tilde{a}(z) = a(z) - [K\nabla\mathcal{H}(0, z)] \cdot \nabla_z c(z). \tag{B.12}$$

Equation (B.11) can be solved if $\tilde{a}(z) = c^2(z)b(z)$, which fixes $c(z)$. As we will see later, the exact form of $c(z)$ will not matter, so we will only assume that this exists. Notice simply that $K\nabla\mathcal{H}(0) = 0$ (since by assumption $z = 0$ is the location of the saddle point on S), and therefore

$$\tilde{a}(0) = a(0), \qquad c(0) = \sqrt{\frac{a(0)}{b(0)}} \tag{B.13}$$

The solution of (B.11) subject to $\lim_{\zeta\to\pm\infty} \phi_1(\zeta) = C_\pm$ is:

$$\phi_1(\zeta) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\zeta} \exp\left(-\frac{1}{2}\zeta'^2\right) d\zeta' - \frac{1}{\sqrt{2}}, \tag{B.14}$$

or, in terms of φ_1

$$\varphi_1(\eta, z) = \sqrt{\frac{\tilde{a}(z)}{\pi b(z)}} \int_{-\infty}^{\eta c(z)} \exp\left(-\frac{1}{2}\frac{\tilde{a}(z)}{b(z)}\eta'^2\right) d\eta' - \frac{1}{\sqrt{2}}. \tag{B.15}$$

We are now in position to find λ_1. To do so multiply both sides of (B.2) by the equilibrium density, $\psi = Z^{-1}e^{-\beta\mathcal{H}}$, and integrate over A_+:

$$\int_{A_+} \left[-(K\nabla\mathcal{H}) \cdot \nabla\varphi_1 + \frac{1}{\beta}K^S : \nabla\nabla\varphi_1\right] \psi du = -\lambda_1 \int_{A_+} \varphi\psi du. \tag{B.16}$$

Since $\nabla\nabla\varphi_1$ is symmetric, $K^A : \nabla\nabla\varphi_1 = 0$, and using Green's Theorem, the left hand side of (B.16) becomes

$$\begin{aligned}\frac{1}{\beta}\int_{A_+} \nabla \cdot \left(\psi K^T \nabla\varphi_1\right) du &= \frac{1}{\beta}\int_S \psi \hat{n}(z) \cdot K^T \nabla\varphi_1(0, z) dz \\ &= \sqrt{\frac{1}{\pi\beta}} Z^{-1} \int_S \sqrt{\frac{\tilde{a}(z)}{b(z)}} \hat{n}(z) \cdot K^T \hat{n}(z) e^{-\beta\mathcal{H}} dz \\ &= \sqrt{\frac{1}{\pi\beta}} Z^{-1} \int_S \sqrt{\tilde{a}(z)b(z)} e^{-\beta\mathcal{H}} dz, \end{aligned} \tag{B.17}$$

where we used $\nabla\varphi_1 = \sqrt{\beta\tilde{a}/\pi b}\hat{n}$ on S and $\hat{n} \cdot K^T\hat{n} = \hat{n} \cdot K\hat{n} = b(z)$. Therefore,

$$\lambda_1 = -\sqrt{\frac{1}{\pi\beta}} \frac{\int_S \sqrt{\tilde{a}(z)b(z)} e^{-\beta\mathcal{H}} d\sigma(u)}{\int_{A_+} \varphi_1 e^{-\beta\mathcal{H}} du}. \tag{B.18}$$

The next step is to evaluate the two integrals in (B.18) using the Laplace method. For large β the denominator is approximated by

$$\begin{aligned}\int_{A_+} \varphi_1(u) e^{-\beta\mathcal{H}} du &= \varphi_1(u_{\min}) \left(\frac{2\pi}{\beta}\right)^{d/2} \frac{1}{\sqrt{\det H_{\min}}} \\ &= \sqrt{\frac{2}{\det H_{\min}}} \left(\frac{2\pi}{\beta}\right)^{d/2}, \end{aligned} \tag{B.19}$$

where $H_{\min} = \nabla\nabla\mathcal{H}(u_{\min})$ is the Hessian at the energy minimum. We also used the fact that in the limit $\beta \to \infty$, $\varphi_1(u_{\min}) \to C_+ = 1/\sqrt{2}$. Evaluating the numerator,

$$\int_S \sqrt{\tilde{a}(z)b(z)}e^{-\beta\mathcal{H}}d\sigma = \left(\frac{2\pi}{\beta}\right)^{(d-1)/2} \sqrt{a(0)b(0)}\frac{1}{\sqrt{\det H_\perp}}, \tag{B.20}$$

where we used (B.13) and $H_\perp$ is the restriction of the Hessian at the saddle point (the origin in our case) to the hyperplane perpendicular to $\hat{n}$. Substituting into (B.18) gives

$$\lambda_1 = -\frac{1}{2\pi}\sqrt{a(0)b(0)}\sqrt{\frac{\det H_{\min}}{\det H_\perp}}e^{-\beta\Delta\mathcal{H}}, \tag{B.21}$$

where $\Delta\mathcal{H} = \mathcal{H}(0,0,0) - \mathcal{H}(\mathbf{u}_{\min})$ is the energy barrier.

B.4. Evaluation of $\hat{n}$, a and b

The direction $\hat{n}_0 = \hat{n}(0)$ can be expressed in terms of the eigenvector, $\hat{t}$, corresponding to the negative eigenvalue λ_- of $K^T H_0$, i.e.

$$K^T H_0 \hat{t} = \lambda_- \hat{t}, \tag{B.22}$$

where $H_0 = \nabla\nabla\mathcal{H}(0)$ is the full Hessian at the origin. The normal vector $\hat{n}_0$ is the unit vector that is perpendicular to the stable manifold of the flow $K\nabla\mathcal{H}$ out of the saddle point, such that $\hat{t}\cdot\hat{n}_0 > 0$. Therefore, if $\hat{v}$ is an eigenvector of KH_0 with positive eigenvalue μ, then $\hat{n}_0 \cdot \hat{v} = 0$. We claim that the vector $K^{-T}\hat{t}$, where $K^{-T} = (K^T)^{-1}$, also has this property and hence, $\hat{n}_0 || K^{-T}\hat{t}$. Indeed, using (B.22), we have

$$\begin{aligned}\lambda_-(K^{-T}\hat{t}\cdot\hat{v}) &= (\lambda_-\hat{t}\cdot K^{-1}\hat{v})\\ &= (K^T H_0\hat{t}\cdot K^{-1}\hat{v})\\ &= (\hat{t}\cdot H_0 K K^{-1}\hat{v}) = (\hat{t}\cdot H_0\hat{v})\\ &= (\hat{t}\cdot K^{-1}KH_0\hat{v})\\ &= (K^{-T}\hat{t}\cdot KH_0\hat{v}) = \mu(K^{-T}\hat{t}\cdot\hat{v}).\end{aligned} \tag{B.23}$$

Therefore,

$$(\mu - \lambda_-)(K^{-T}\hat{t}\cdot\hat{v}) = 0, \tag{B.24}$$

and since $\lambda_- < 0 < \mu$, we conclude that for any eigenvector $\hat{v}$ of KH_0 with positive eigenvalue μ. This implies

$$\hat{n}_0 = \pm \frac{1}{|K^{-T}\hat{t}|} K^{-T}\hat{t}. \tag{B.25}$$

We also have,

$$H_0\hat{t} \cdot \hat{v} = K^{-T}K^T H_0\hat{t} \cdot \hat{v} = \lambda_- K^{-T}\hat{t} \cdot \hat{v} = \pm|K^{-T}\hat{t}|\lambda_- \hat{n}_0 \cdot \hat{v} = 0, \tag{B.26}$$

which implies that,

$$\hat{n}_0 = \mp \frac{1}{|H_0\hat{t}|} H_0\hat{t}. \tag{B.27}$$

The sign in (B.27), which is always opposite to that of (B.25), is determined by comparing the two equations for $\hat{n}_0$. Substituting into (B.9) yields

$$a(0) = |\lambda_-|, \qquad b(0) = |\lambda_-| \frac{|\hat{t} \cdot \hat{n}_0|}{|H_0\hat{t}|}. \tag{B.28}$$

B.5. Evaluation of $H_\perp$

We now show that

$$\det H_\perp = \frac{|\det H_0| \left|\hat{t} \cdot \hat{n}_0\right|}{\left|H_0\hat{t}\right|}. \tag{B.29}$$

Denote

$$H_m = H_0 + m\hat{n}_0 \otimes \hat{n}_0^T. \tag{B.30}$$

The first step is to show that for large enough m, H_m is positive definite. H_0 is symmetric and therefore diagonalizable. It has one unstable and $d-1$ stable directions. Denoting the eigenvector of H_0 corresponding to the negative eigenvalue by u_1, and the rest by $u_2, \ldots, u_d$, the definition of $\hat{n}_0$ implies it is not perpendicular to u_1. H_m is also symmetric and hence diagonalizable. Denoting its eigenvectors by $v_1, \ldots, v_d$, we have $v_i \cdot H_m v_i = v_i \cdot H_0 v_i + m(v_i \cdot \hat{n}_0)^2$. If $v_i \perp \hat{n}_0$, then it can be written as a linear combination of $u_2, \ldots, u_d$, and $v_i \cdot H_m v_i = v_i \cdot H_0 > 0$. Otherwise, $v_i \cdot H_m v_i$ is positive for large enough m.

The next step is to express $\det H_\perp$, which is also positive definite, using Gaussian integrals:

$$\begin{aligned}\pi^{(d-1)/2} (\det H_\perp)^{-1/2} &= \int_{y\cdot\hat{n}=0} e^{-y\cdot H_\perp y} d\sigma \\ &= \int_{y\cdot\hat{n}=0} e^{-y\cdot H_0 y} d\sigma = \int_{y\cdot\hat{n}=0} e^{-y\cdot H_m y} d\sigma. \end{aligned} \tag{B.31}$$

Also,

$$\pi^{(d-1)/2}\left(\det H_m\right)^{-1/2} = \int_{\mathbb{R}^d} e^{-u\cdot H_m u} du$$
$$= (\hat{n}_0 \cdot \hat{t}) \int_{y\cdot\hat{n}=0} d\sigma \int_{\mathbb{R}} ds e^{-(y+s\hat{t})\cdot H_m (y+s\hat{t})} \qquad \text{(B.32)}$$

where we used a change of variables $u = y + s\hat{t}$ with $y \perp \hat{n}_0$ whose Jacobian is $\hat{n}_0 \cdot \hat{t}$. Using (B.27) this integral becomes

$$\pi^{(d-1)/2}\left(\det H_m\right)^{-1/2} = (\hat{n}_0 \cdot \hat{t}) \int_{y\cdot\hat{n}=0} e^{-y\cdot H_m y} d\sigma \int_{\mathbb{R}} ds e^{-s^2[|H_0\hat{t}||\hat{n}_0\cdot\hat{t}|+m(\hat{n}_0\cdot\hat{t})^2]}$$
$$= \pi^{(d-1)/2}\left(\det H_\perp\right)^{-1/2}\left(|H_0\hat{t}||\hat{n}_0\cdot\hat{t}| + m|\hat{n}_0\cdot\hat{t}|^2\right)^{-1/2}. \qquad \text{(B.33)}$$

Hence,

$$|\det H_\perp| = |\det H_m| \frac{|\hat{n}_0 \cdot \hat{t}|}{|H_0\hat{t}| + m|\hat{n}_0 \cdot \hat{t}|}. \qquad \text{(B.34)}$$

The last step is relating $\det H_m$ to $\det H_0$. Using (B.27),

$$H_m = H_0 \left(Id - \frac{m}{|H_0\hat{t}|} \hat{t} \otimes \hat{n}_0^T \right), \qquad \text{(B.35)}$$

where Id denotes the $d \times d$ identity matrix. The matrix in parenthesis has d eigenvectors: $d-1$ vectors perpendicular to $\hat{n}_0$ with eigenvalue 1 and $\hat{t}$ with eigenvalue $1 - m|\hat{n}_0 \cdot \hat{t}|/|H_0\hat{t}|$. Hence,

$$\det H_m = \left(1 - \frac{m|\hat{n}_0 \cdot \hat{t}|}{|H_0\hat{t}|}\right) \det H_0. \qquad \text{(B.36)}$$

Substituting into (B.34) and taking the limit $m \to \infty$ yields the formula for the restricted Hessian, (B.29).

B.6 Conclusion

Substituting (B.28) and (B.36) into (B.21) yields,

$$\lambda_1 = -\frac{|\lambda_-|}{2\pi} \sqrt{\left| \frac{\det H_{\min}}{\det H_0} \right|} e^{-\beta \mathcal{H}}. \qquad \text{(B.37)}$$

Note that although the algebra is more complicated, the final formula is exactly the same as the well known Kramers formula, obtained in the simple case of $K = Id$.(4,13,28)

In our model the effective dynamics is given by (B.1). The limiting equation is three dimensional with parameters according to (B.12). Substituting into (B.37) yields

$$\lambda_1 = -\frac{|\lambda_-|}{\sqrt{2}\pi} e^{-\beta}, \tag{B.38}$$

where λ_- is the negative root of the characteristic polynomial of $K^T H_0$:

$$\lambda_-^3 - \lambda_-^2 - (4-\gamma)\lambda_- + 4 = 0. \tag{B.39}$$

Pollak *et al.* [(34)] obtain the same transition rate for the limiting equations from a different approach. Using the known rate obtained by the VTST method at fixed N they construct an eigenfunction for the forward (Fokker-Planck) operator. Our method is independent of the details of the Kac-Zwanzig model itself and only makes use of the limiting process. In addition, solving the eigenvalue problem shows there is only a single eigenvalue of order $e^{-\beta}$.

ACKNOWLEDGMENTS

We thank Alexandre Chorin, Matthias Heymann, Jose Koiller and Paul Wright for useful discussions and suggestions. We are especially grateful to Ray Kapral for sharing his insight on the Kac-Zwanzig model and to Zeev Schuss for introducing us to the work of Pollak et al and sharing some unpublished results. Partial support from NSF through Grants DMS01-01439, DMS02-09959 and DMS02-39625, and from ONR through Grant N-00014-04-1-0565 is gratefully acknowledged.

REFERENCES

1. G. Ariel and E. Vanden-Eijnden, in preparation.
2. C. H. Bennett, Molecular dynamics and transition state theory: the simulation of infrequent events. In R. E. Christoffersen (ed.), *Algorithms for chemical computations*, pp. 63–97 (Amer. Chem. Soc., Washington D.C., 1977).
3. P. G. Bolhuis, C. Dellago, D. Chandler and P. L. Geissler. Transition path sampling: Throwing ropes over mountain passes, in the dark. *Ann. Rev. Phys. Chem.* **53**:291–318 (2002).
4. A. Bovier, M. Eckhoff, V. Gayrard and M. Klein, Metastability in reversible diffusion processes—I. Sharp asymptotics for capacities and exit times. *J. Euro. Math. Soc.* **6**:399–424 (2004).
5. B. Cano and A. M. Stuart, Underresolved simulations of heat baths. *J. Comp. Phys.* **169**:193–214 (2001).
6. D. Chandler, Statistical mechanics of isomerization dynamics in liquids and the transition state approximation, *J. Chem. Phys.* **68**:2959–2970 (1978).
7. J. R. Chaudhuri, S. K. Banik, B. C. Bag and D. S. Ray, Analytical and numerical investigation of escape rate for a noise driven bath. *Phys. Rev. E* **63**:Art. 061111 (2001).
8. C. Dellago, P. G. Bolhuis and P. L. Geissler, Transition path sampling. *Adv. Chem. Phys.* **123**:1–78 (2002).

9. W. E and E. Vanden-Eijnden, Metastability, conformation dynamics, and transition pathways in complex systems. In S. Attinger and P. Koumoutsakov (eds.), *Lecture notes in computational science and engineering* **39**:35–68 (Springer, Berlin, 2004).
10. W. E, W. Ren and E. Vanden-Eijnden, Finite temperature string method for the study of rare events. *J. Phys. Chem. B* **109**:6688–6693 (2005).
11. W. E, W. Ren and E. Vanden-Eijnden, Transition pathways in complex systems: Reaction coordinates, isocommittor surfaces, and transition tubes. *Chem. Phys. Lett.* **413**:242–247 (2005).
12. W. E and E. Vanden-Eijnden, Toward a theory of transitions paths. *J. Stat. Phys.* **123**:503–523 (2006)
13. H. Eyring, *J. Chem. Phys.* **3**:107 (1935).
14. G. W. Ford, M. Kac and P. Mazur, Statistical mechanics of assemblies of coupled oscillators. *J. Math. Phys.* **6**:504–515 (1965).
15. G. W. Ford and M. Kac, On the Quantum Langevin Equation. *J. Stat. Phys.* **46**:803–810 (1987).
16. G. R. Fleming and G. Wolynes, Chemical dynamics in solution. *Physics Today* **43**:36–43 (1990).
17. D. Frenkel and B. Smit, *Understanding Molecular Dynamics*, Academic Press, San Diego (1996).
18. C. W. Gardiner, *Handbook of Stochastic Methods*, 2nd edn., Springer, Berlin (1985).
19. D. Givon, R. Kupferman and A. M. Stuart, Extracting macroscopic dynamics: Model problems and algorithms. *Nonlinearity* **17**:R55–R127 (2004).
20. H. Grabert, Escape from a metastable well: The Kramers turnover problem. *Phys. Rev. Lett.* **61**:1683–1686 (1988).
21. O. H. Hald and R. Kupferman, Asymptotic and numerical analyses for mechanical models of heat baths. *J. Stat. Phys.* **106**:1121–1184 (2002).
22. J. Horiuti, *Bull. Chem. Soc. Jpn.* **13**:210 (1938).
23. W. Huisinga, C. Schutte and A. M. Stuart, Extracting macroscopic stochastic dynamics: Model problems. *Comm. Pure Appl. Math.* **56**:0234 (2003).
24. J. C. Keck, *Disc. Faraday Soc.* **33**:173 (1962).
25. R. Kupferman, Fractional Kinetics in Kac-Zwanzig heat bath models. *J. Stat. Phys.* **111**:291–326 (2004).
26. R. Kupferman and A. M. Stuart, Fitting SDE models to nonlinear Kac-Zwanzig heat bath models. *Phys. D-Nonlinear phenomena* **199**:279–316 (2004).
27. R. Kupferman, A. M. Stuart, J. R. Terry and P. F. Tupper, Long term behavior of large mechanical systems with random initial data. *Stoc. and Dyn.* **2**:533–562 (2002).
28. R. S. Maier and D. L. Stein, Limiting exit location distributions in the stochastic exit problem. *SIAM J. Appl. Math.* **57**:752–790 (1997).
29. B. J. Matkowsky and Z. Schuss, The exit problem for randomly perturbed dynamical systems. *SIAM J. App. Math.* **33**:365–382 (1977).
30. P. Pechukas, *Ann. Rev. Phys. Chem.* **32**:159–177 (1981).
31. E. Pollak, H. Grabert and P. Hänggi, Theory of activated rate processes for arbitrary frequency dependent friction: Solution of the turnover problem. *J. Chem. Phys.* **91**:4073–4087 (1989).
32. E. Pollak, S. C. Tucker and B. H. Berne, Variational transition-state theory for reaction rates in dissipative systems. *Phys. Rev. Lett.* **65**:1399–1402 (1990).
33. E. Pollak and P. Talkner, Activated rate processes: Finite-barrier expansion for the rate in the spatial-diffusion limit. *Phys. Rev. E* **47**:922–933 (1993).
34. E. Pollak, A. M. Berezhkovskii and Z. Schuss, Activated rate processes: A relation between Hamiltonian and stochastic theories. *J. Chem. Phys.* **100**:334–339 (1994).
35. W. Ren, E. Vanden-Eijnden, P. Maragakis and E. Weinan, Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide. *J. Chem. Phys.* **123**:134109 (2005).

36. A. M. Stuart and J. O. Warren, Analysis and experiments for a computational model of a heat bath. *J. Stat. Phys.* **97**:687–723 (1999).
37. Z. Schuss and B. J. Matkowsky, The exit problem: A new approach to diffusion across potential barriers. *SIAM J. App. Math.* **36**:604–623 (1979).
38. Z. Schuss, Singular perturbation methods on stochastic differential equations of mathematical physics. *SIAM Rev.* **22**:119–155 (1980).
39. C. Schütte and W. Huisinga, Biomolecular Conformations as metastable sets of Markov chains, *Proceedings of the Thirty-Eighth Annual Allerton Conference on Communication, Control, and Computing, Monticello, Illinois*: 1106–1115 (2000).
40. F. A. Tal and E. Vanden-Eijnden, Transition state theory and dynamical corrections in ergodic systems. *Nonlinearity* **19**:501–509 (2006).
41. D. G. Truhlar and B. C. Garett, *Ann. Rev. Phys. Chem.* **35**:159–189 (1984).
42. M. Tuckerman, B. J. Berne and G.J. Martina, Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **97**:1990–2001 (1992).
43. T. Uzer, C. Jaffé, J. Palacian, P. Yanguas and S. Wiggins, The geometry of reaction dynamics. *Nonlinearity* **15**:957–992 (2002).
44. E. Vanden-Eijnden and F. Tal, Transition state theory: Variational formulation, dynamical corrections, and error estimates. *J. Chem. Phys.* **123**:184103 (2005).
45. L. Verlet, Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **159**:98–103 (1967).
46. H. Waalkens and S. Wiggins, Direct construction of a dividing surface of minimal flux for multi-degree-of-freedom systems that cannot be recrossed. *J. Phys. A: Math. Gen.* **37**:L435–L445 (2004).
47. E. Wigner, *Trans. Faraday Soc.* **34**:29 (1938).
48. R. Zwanzig, Nonlinear generalized langevin equations. *J. Stat. Phys.* **9**:215–220 (1973).